Deblending galaxies with variational autoencoders

Bastien Arcelin

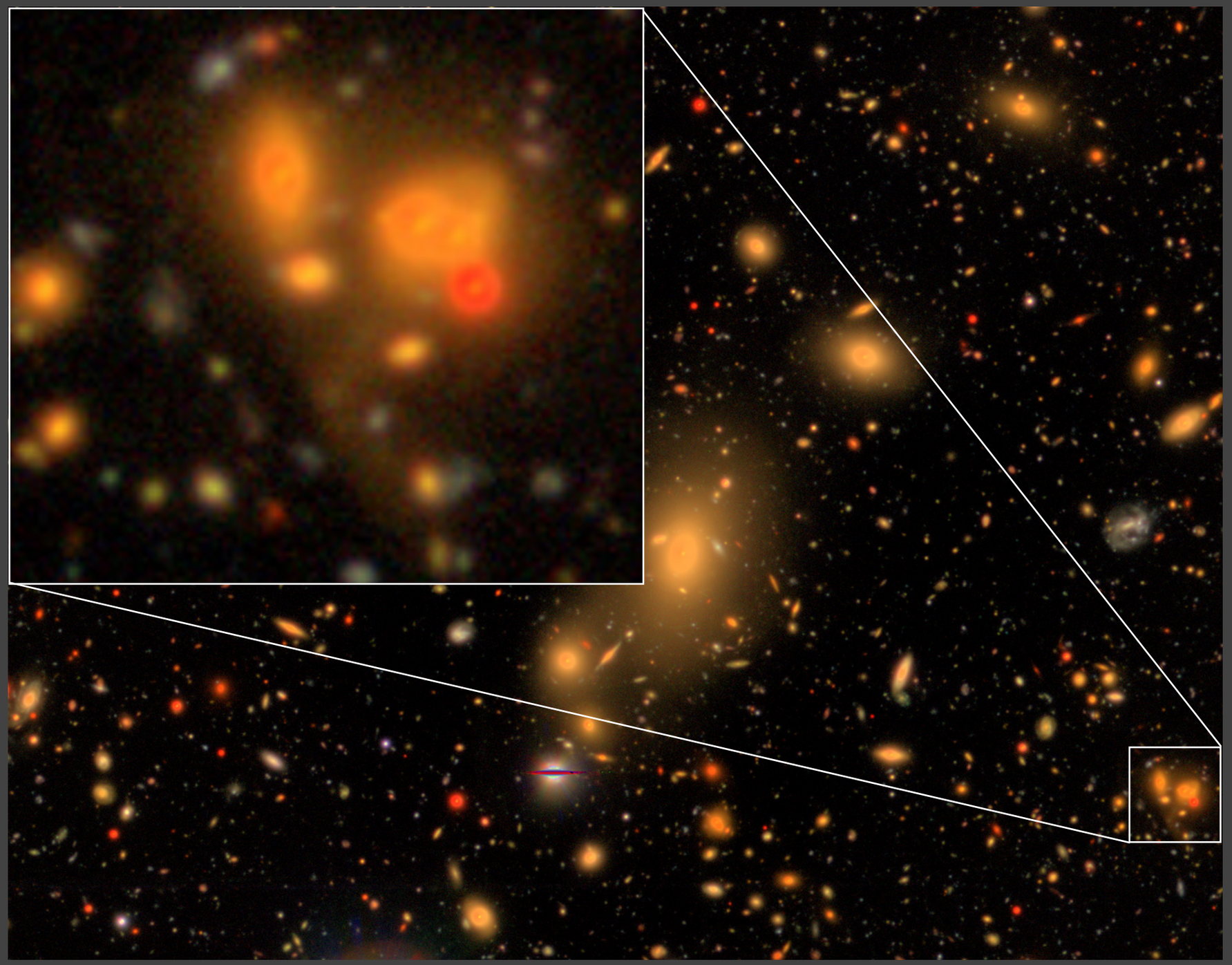
Co-authors: Cyrille Doux, Cécile Roucelle, Eric Aubourg

LSST data

Blending

- Will look a lot like HSC data
- HSC: 58% of the detected objects are identified as blended*
- Systematic in shear measurement

* Bosch et al. (2017)



HSC image of small piece of COSMOS field (https://www.naoj.org/Topics/2017/02/27/index.html)

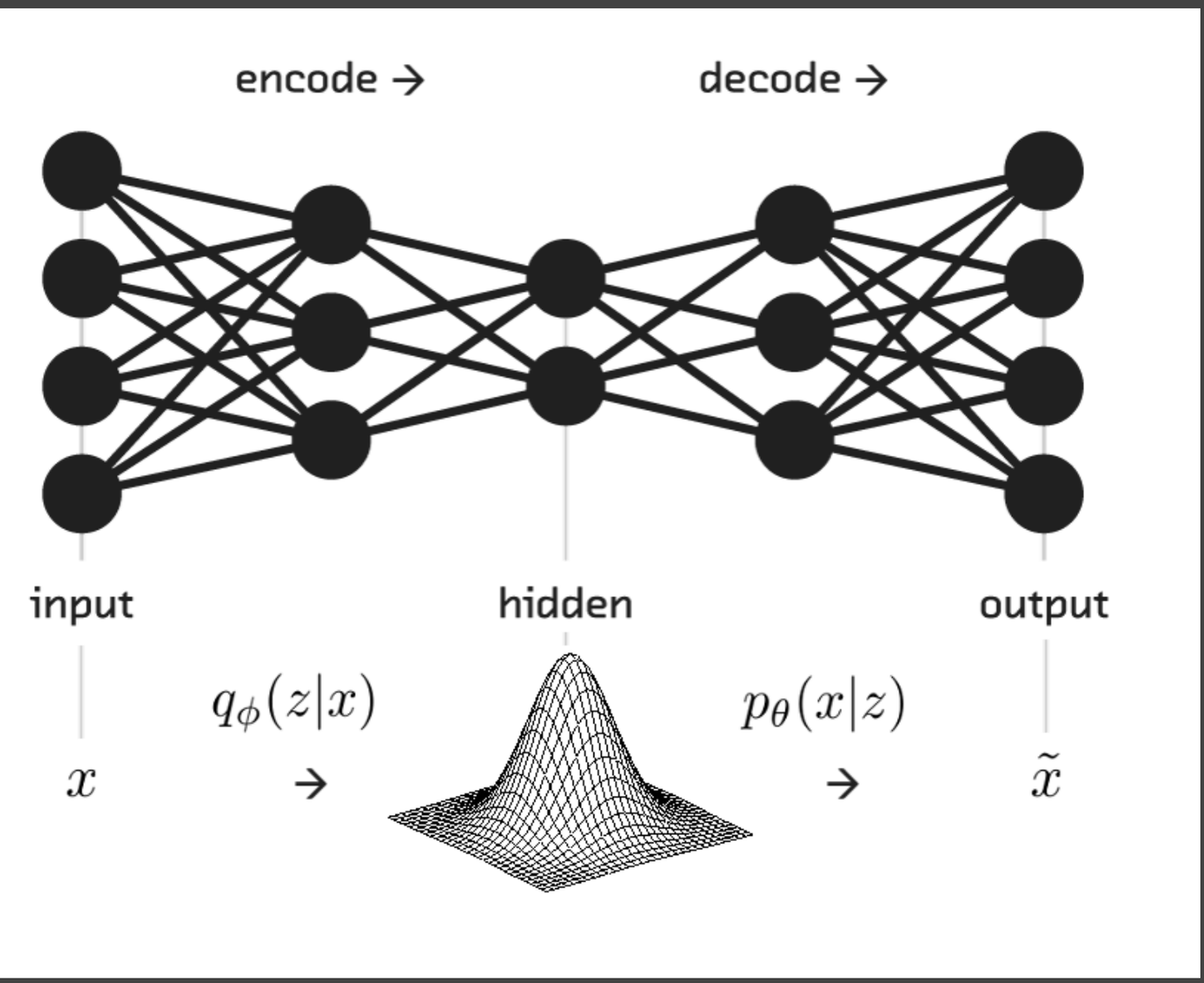
Goals and motivation

Which algorithm ? Which parameters ?

- Create a model for galaxy images from the data → Generative model
- Probabilistic approach → VAE
- Accurate reproduction of relevant weak lensing parameters → Shape and flux reproduction

Variational AutoEncoder (VAE)

(Kingma +2014)



- http://blog.fastforwardlabs.com

Loss function :

$$ELBO = -D_{KL}\left(q_\phi(z|x) \parallel p(z)\right) + \mathbb{E}(\log p_\theta(x|z))$$

Encoder — Regularization term

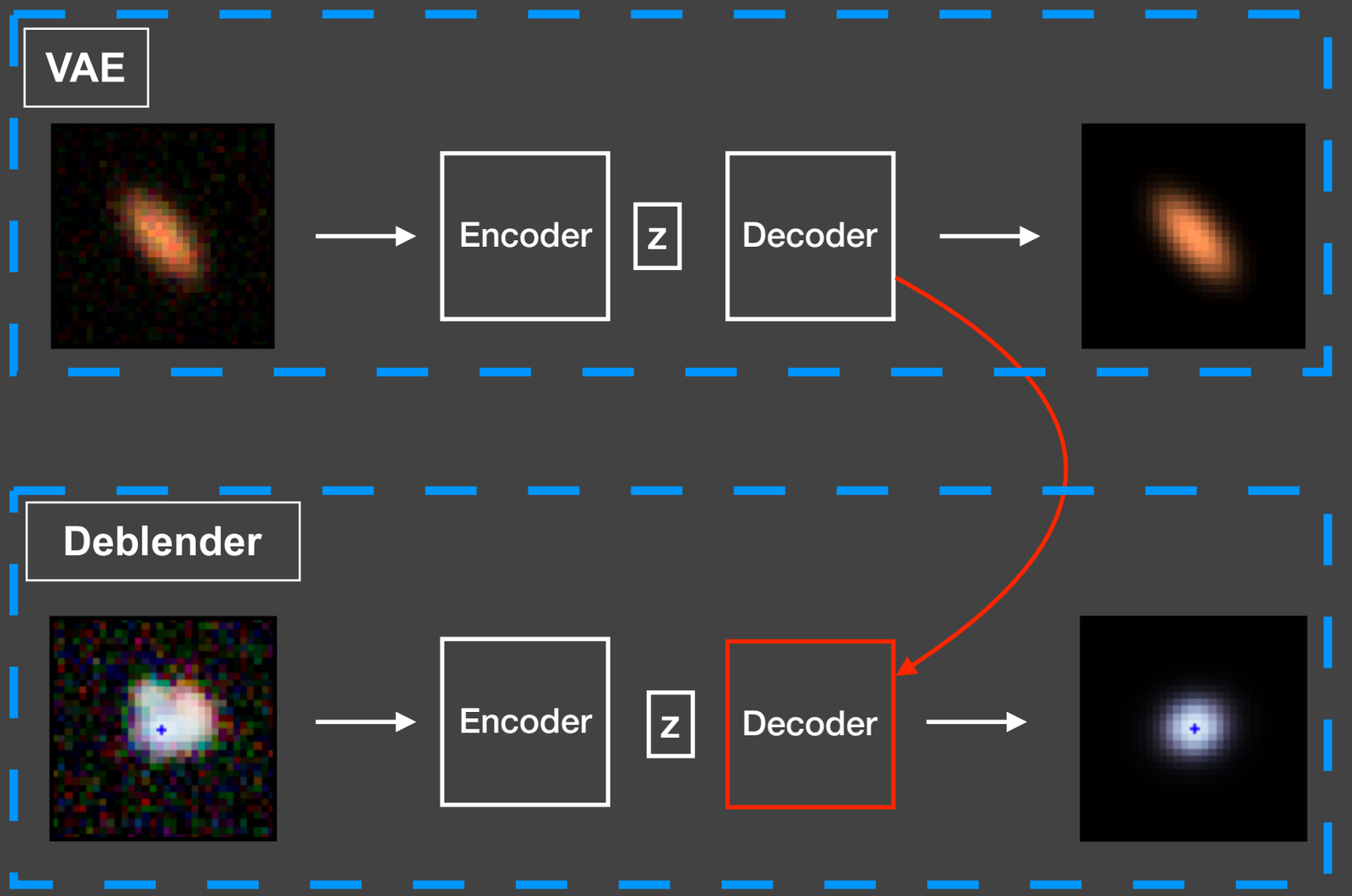
Decoder — Reconstruction term

Machine learning for deblending

Two neural networks:

- VAE (*Kingma+2014*) :
  - Learn a latent variable (z) generative model p(X|z)
  - Approximate the posterior p(z|X) with an encoder

- Deblender:
  - Use fixed generative model from VAE
  - Train a new network that learns to approximate $p(z_{center}|X_{blended})$
  - ~ Perform deblending in latent space



LSST+ Euclid data

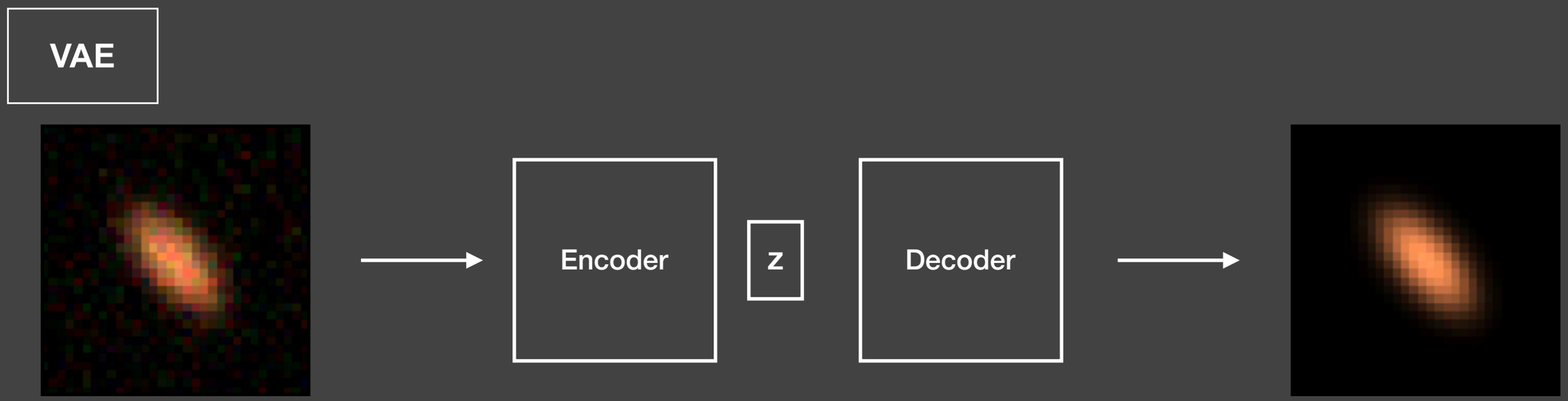
Why using Euclid data:

- ✓ Adding infrared bands (x3)
- ✓ Adding a wide optical band
- ✓ Better resolution (no atmospheric PSF)

DES data - Peter Melchior’s slides

CLASH WFC3/IR data - Peter Melchior’s slides

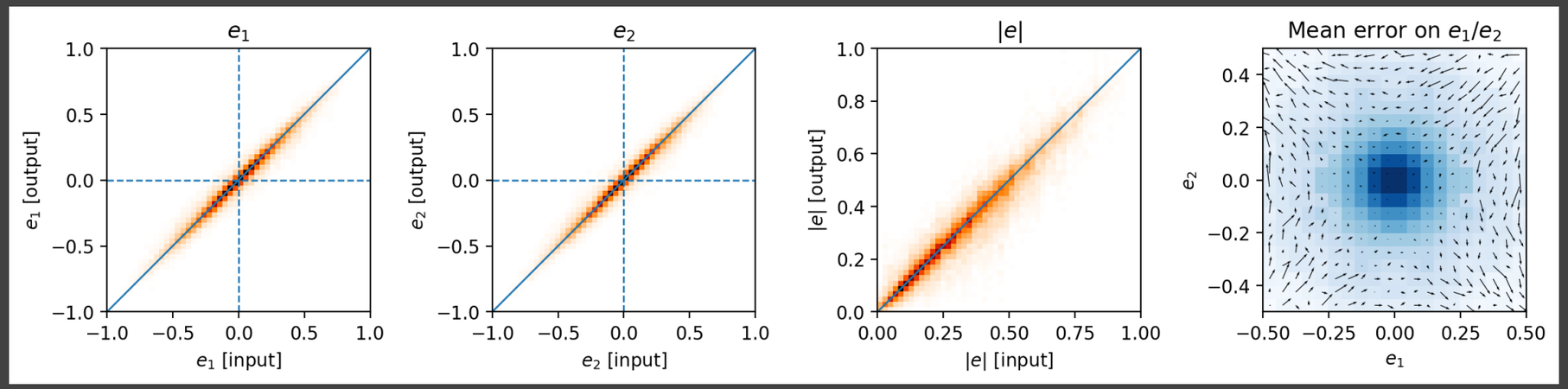
VAE results

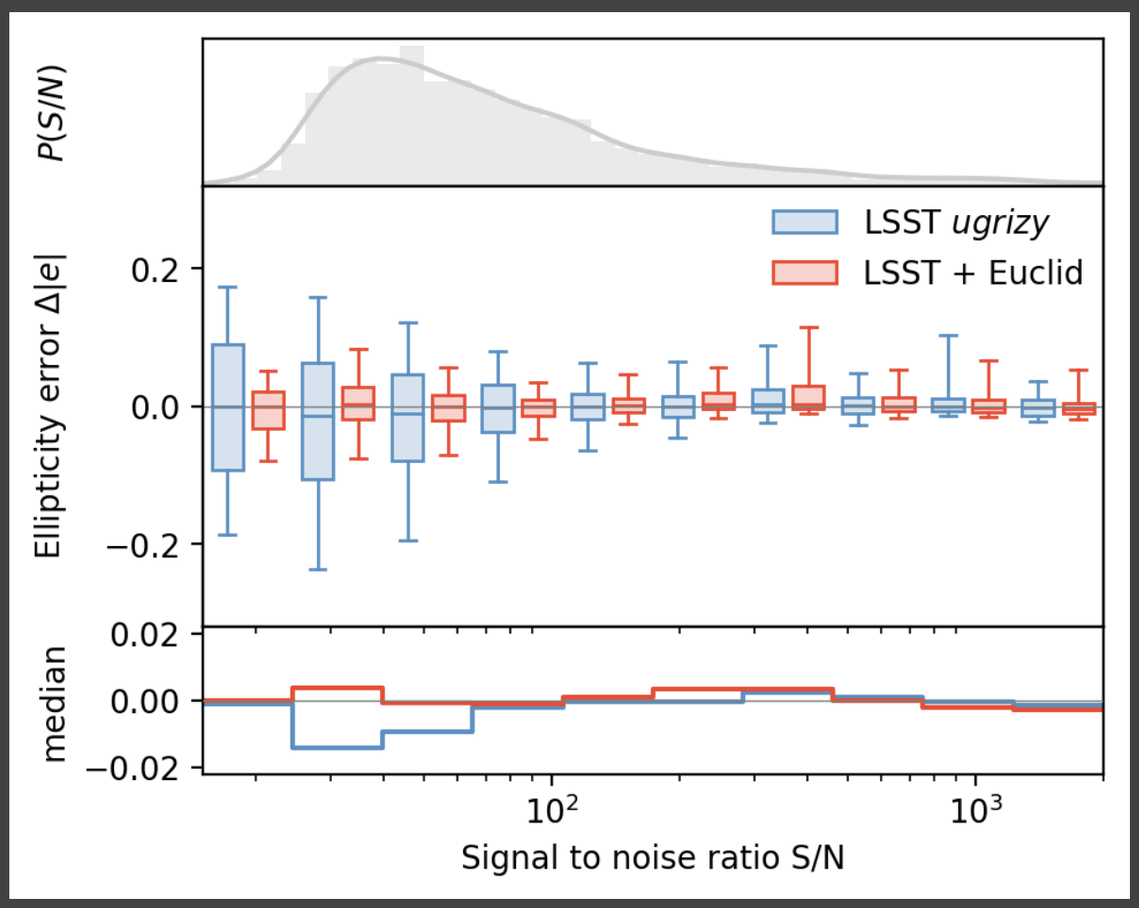


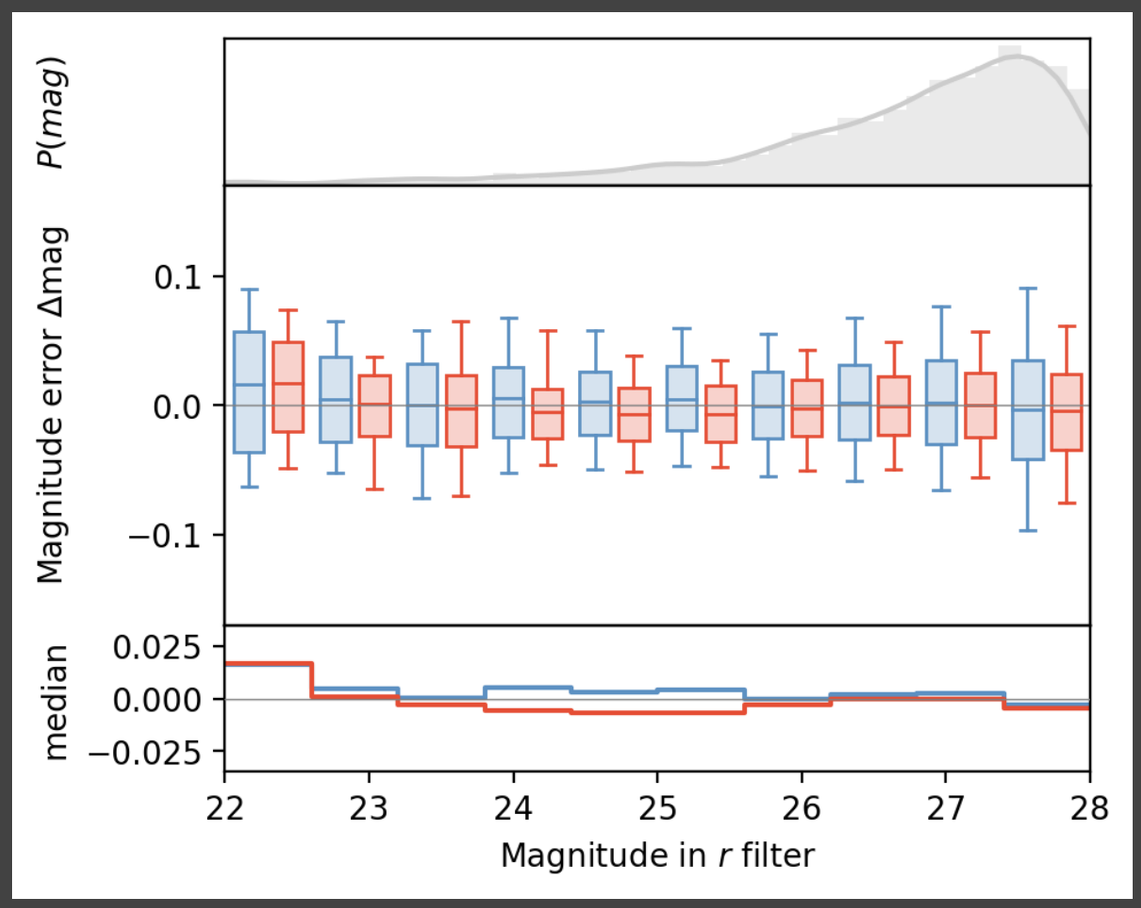
VAE results

Shear and magnitude reproduction

LSST

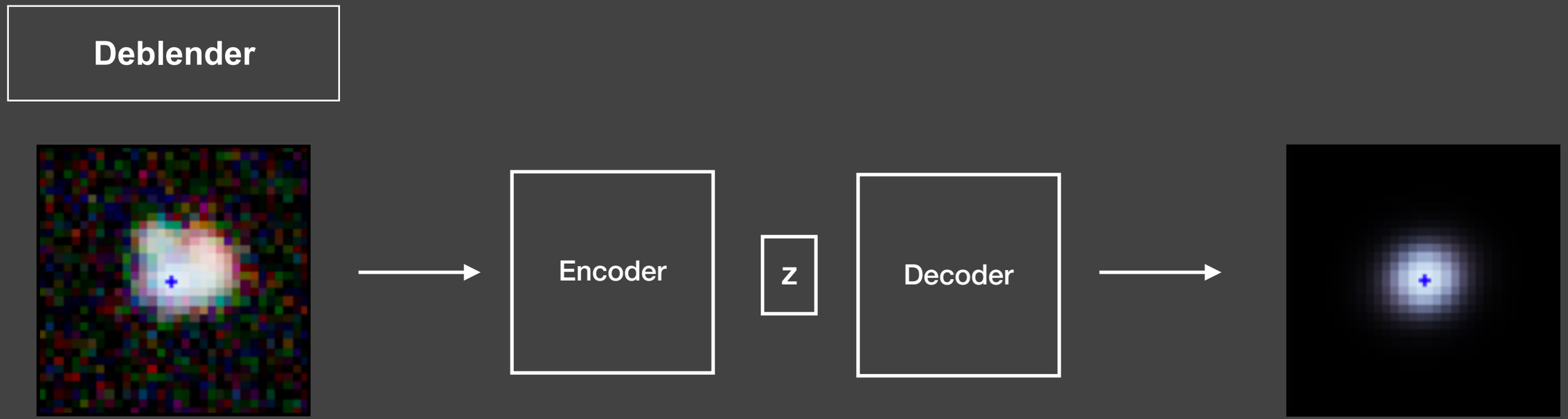






Arcelin et al. (2020)

Deblender results



Blending metric:

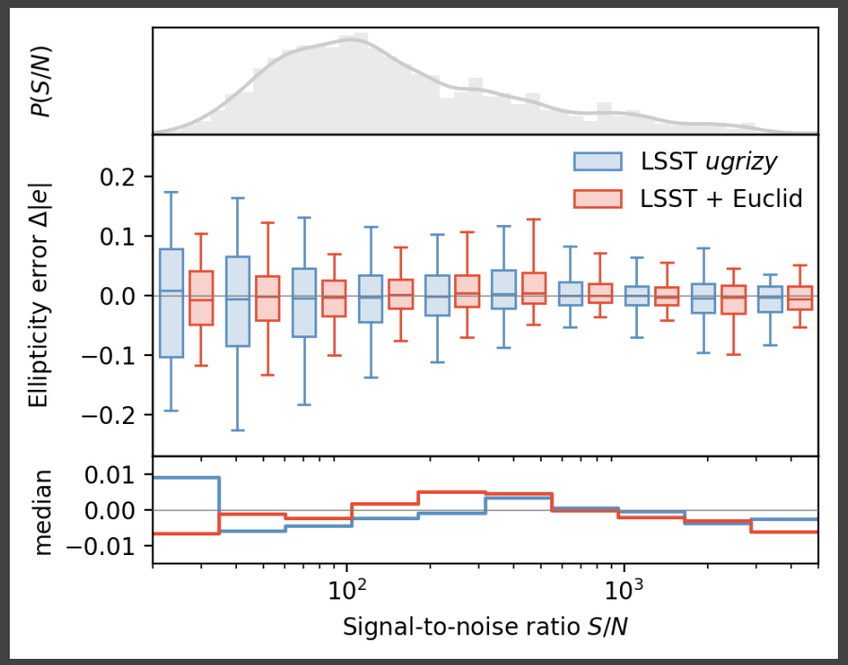
- ***Total blend rate (SCARLET, Melchior et al. (2018)) :***

$$B_{tot} = 1 - \frac{< I_{centered}, I_{centered} >}{< I_{centered}, I_{total} >} \qquad < I_p, I'_p > = \sum_{p \in \{\text{pixels}\}} I_p I'_p$$

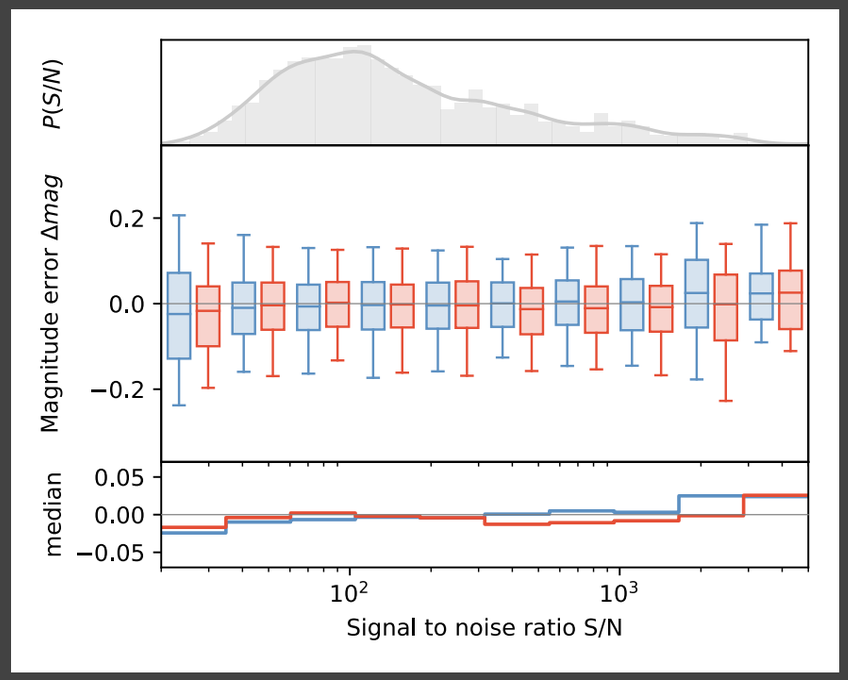
Deblender results

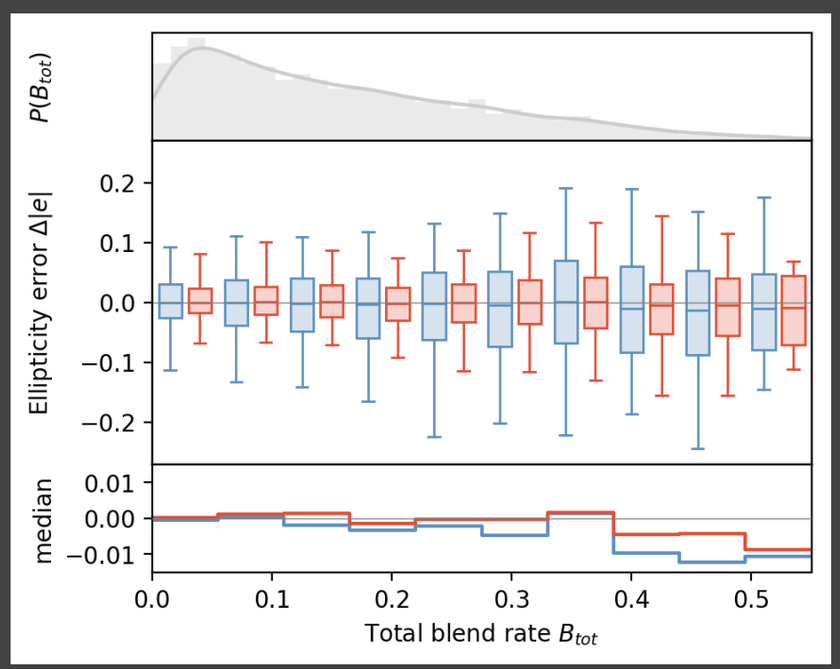
Shear and magnitude reproduction

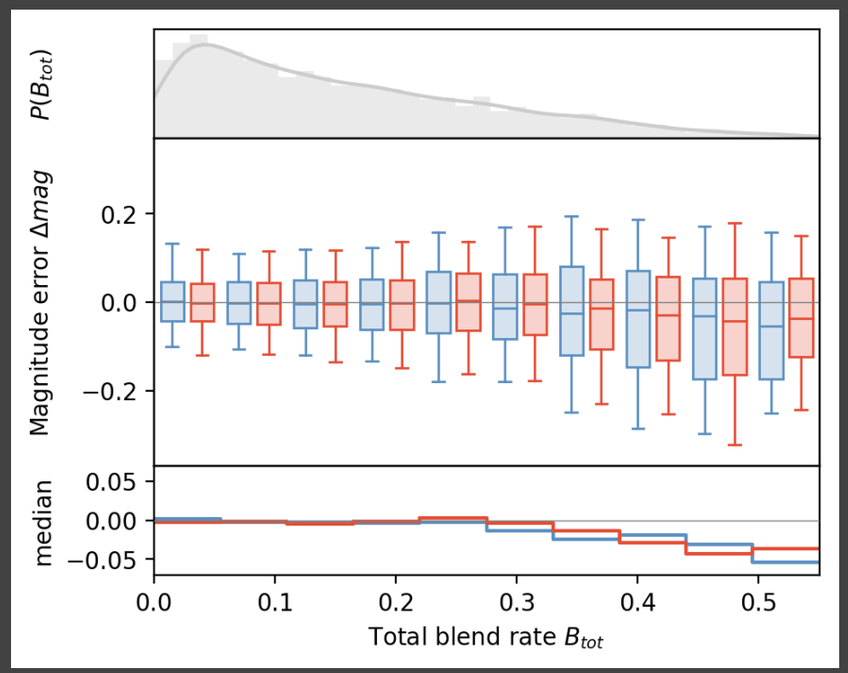
Ellipticity error



r-band magnitude error





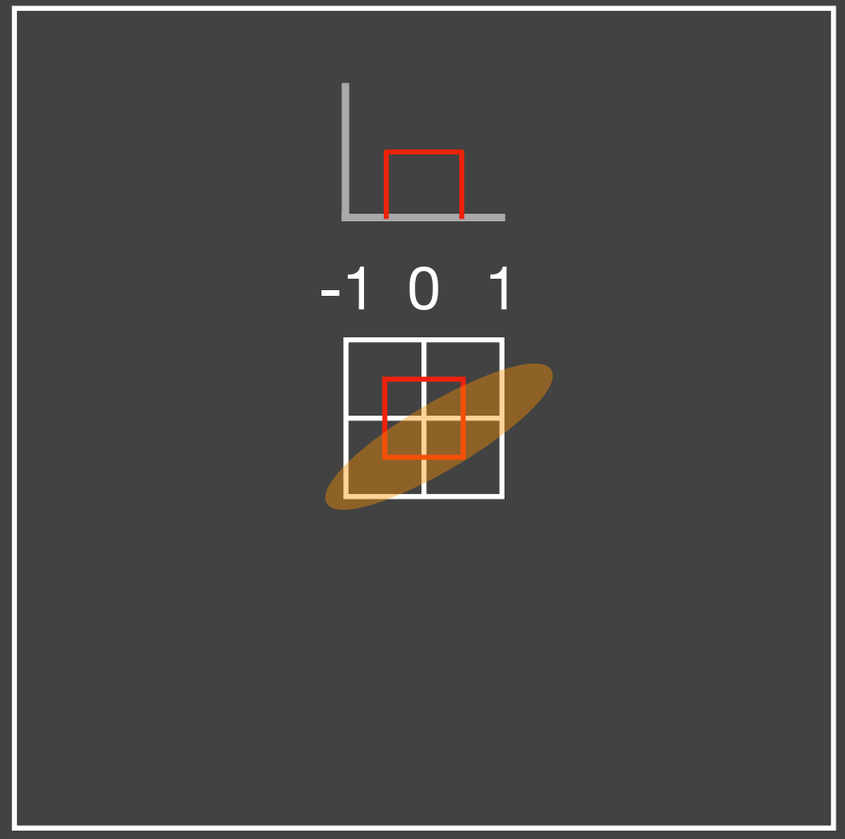


Arcelin et al. (2020)

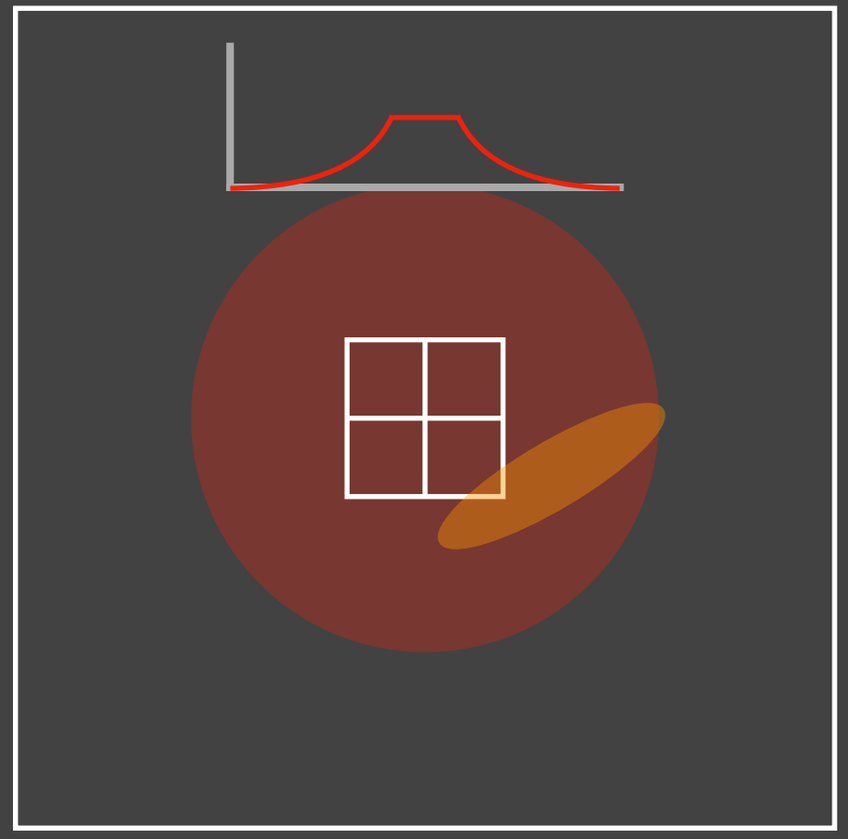
Decentering

In reality, it is impossible to have a perfectly centered galaxy on the stamp: 2 tests configurations

First configuration:
Pixelisation decentering



Second configuration:
Detection algorithm decentering

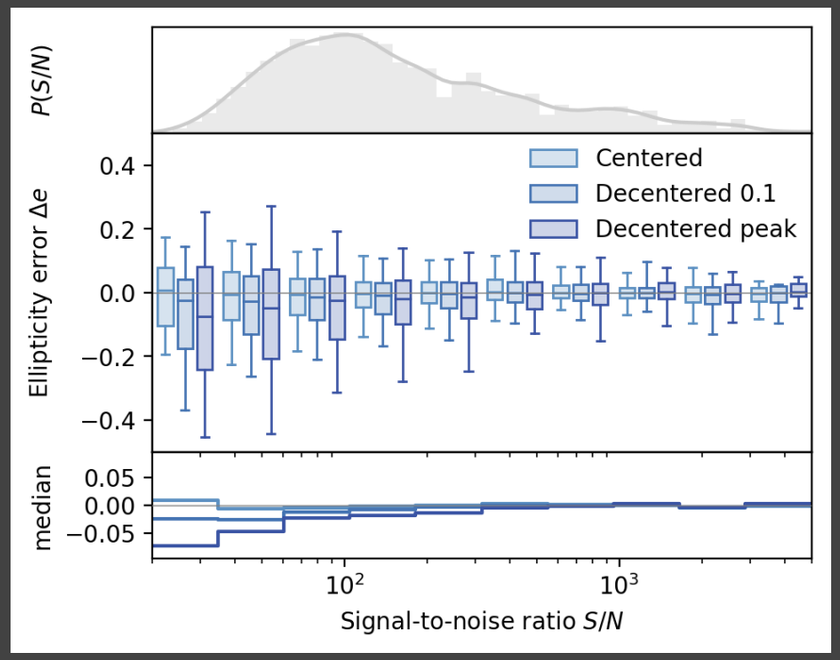


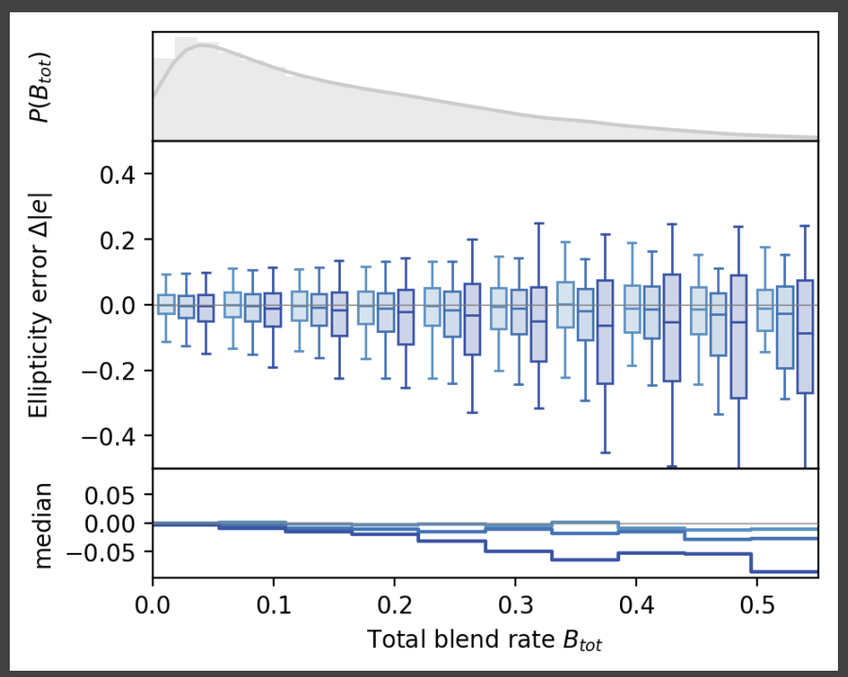
Decentering results

Shear and magnitude reproduction

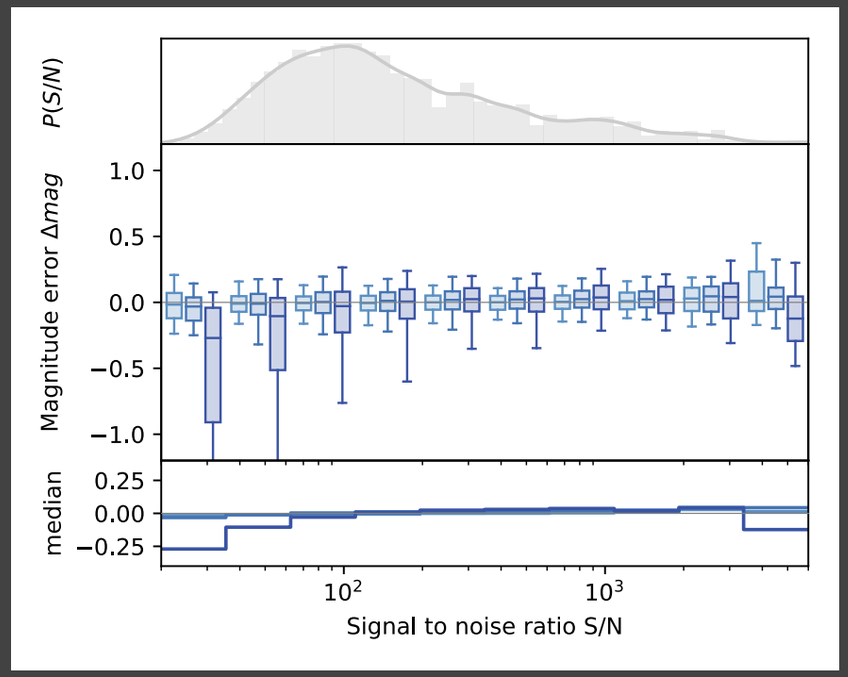
LSST

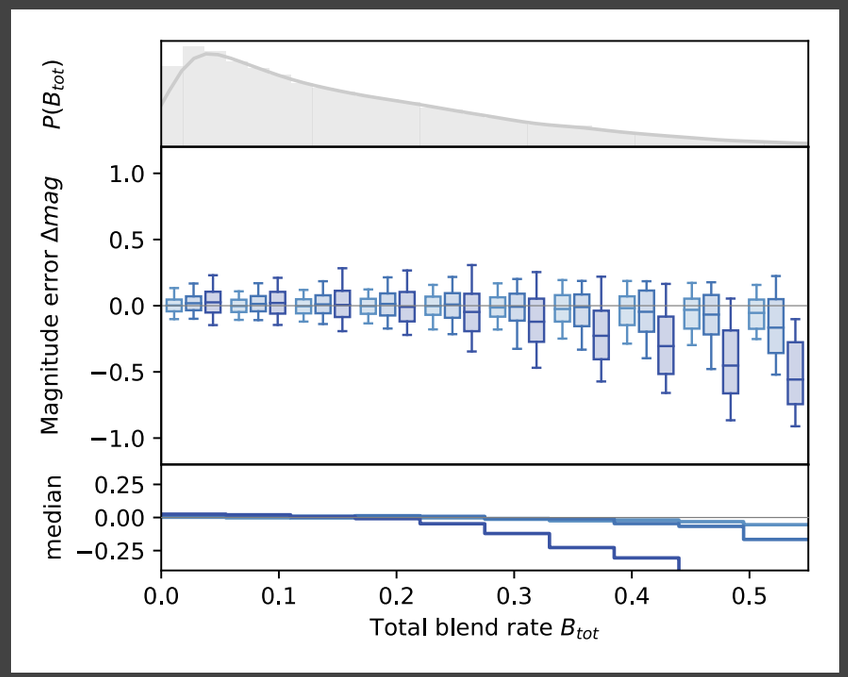
Ellipticity error





r-band magnitude error





Arcelin et al. (2020)

Real images

It works on simulation but what about real images ?

Real images

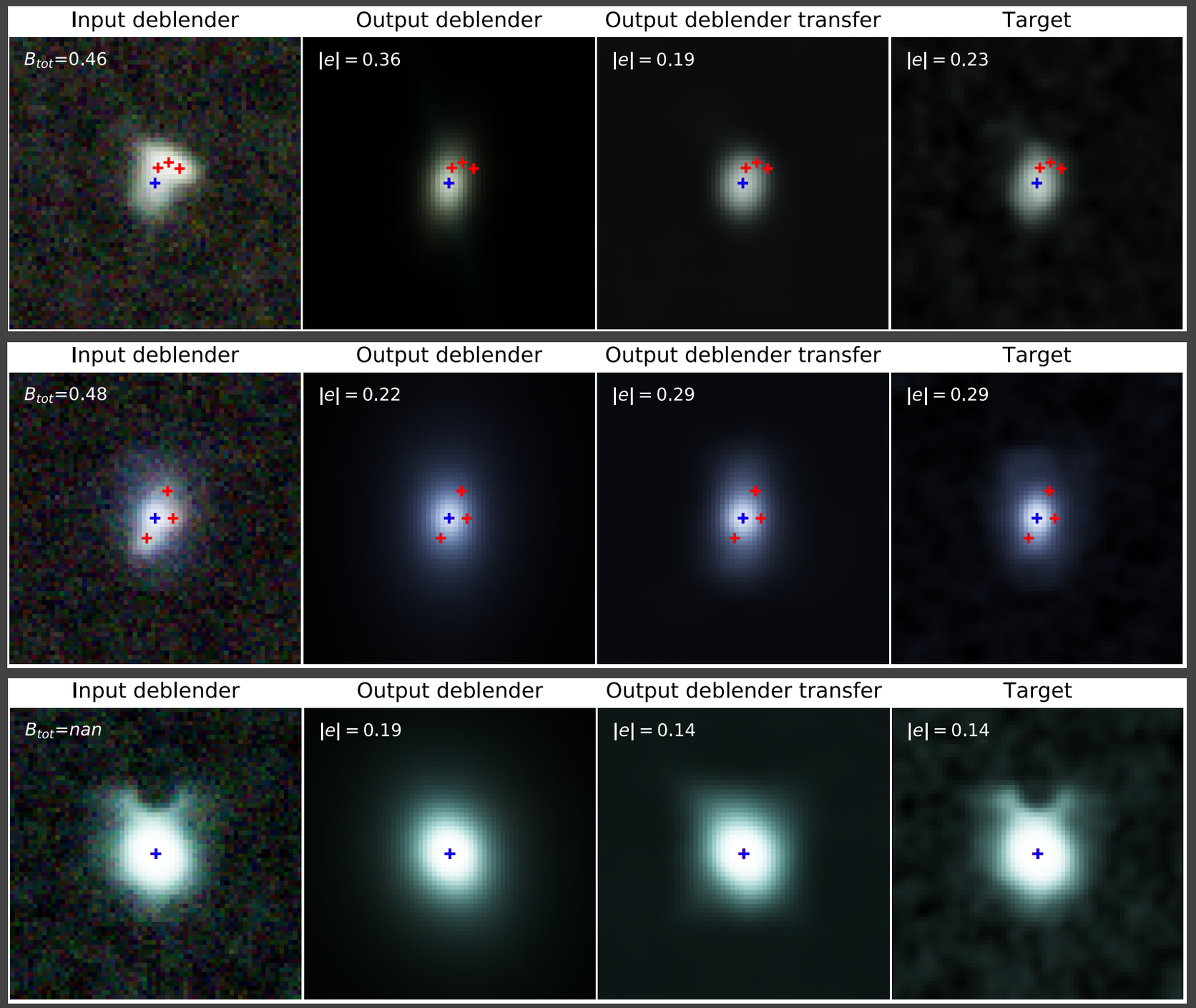
Transfer learning

Transfer learning:

- 80% of simulated data
- 20% of real data

Difficult to have a clean sample:

- Clear blends
- Residual of image processing
- Correlated noise in real images



Arcelin et al. (2020)

Conclusion

- VAE/Deblender:
  Learn/recover accurately features of single galaxy images

- Combining ground and space data:
  Significant improvement in shape reconstruction

- Decentering:
  Our method is quite robust to decentering

- Real images:
  First test with transfer learning. Difficulty to have a clean sample of individual galaxy images.

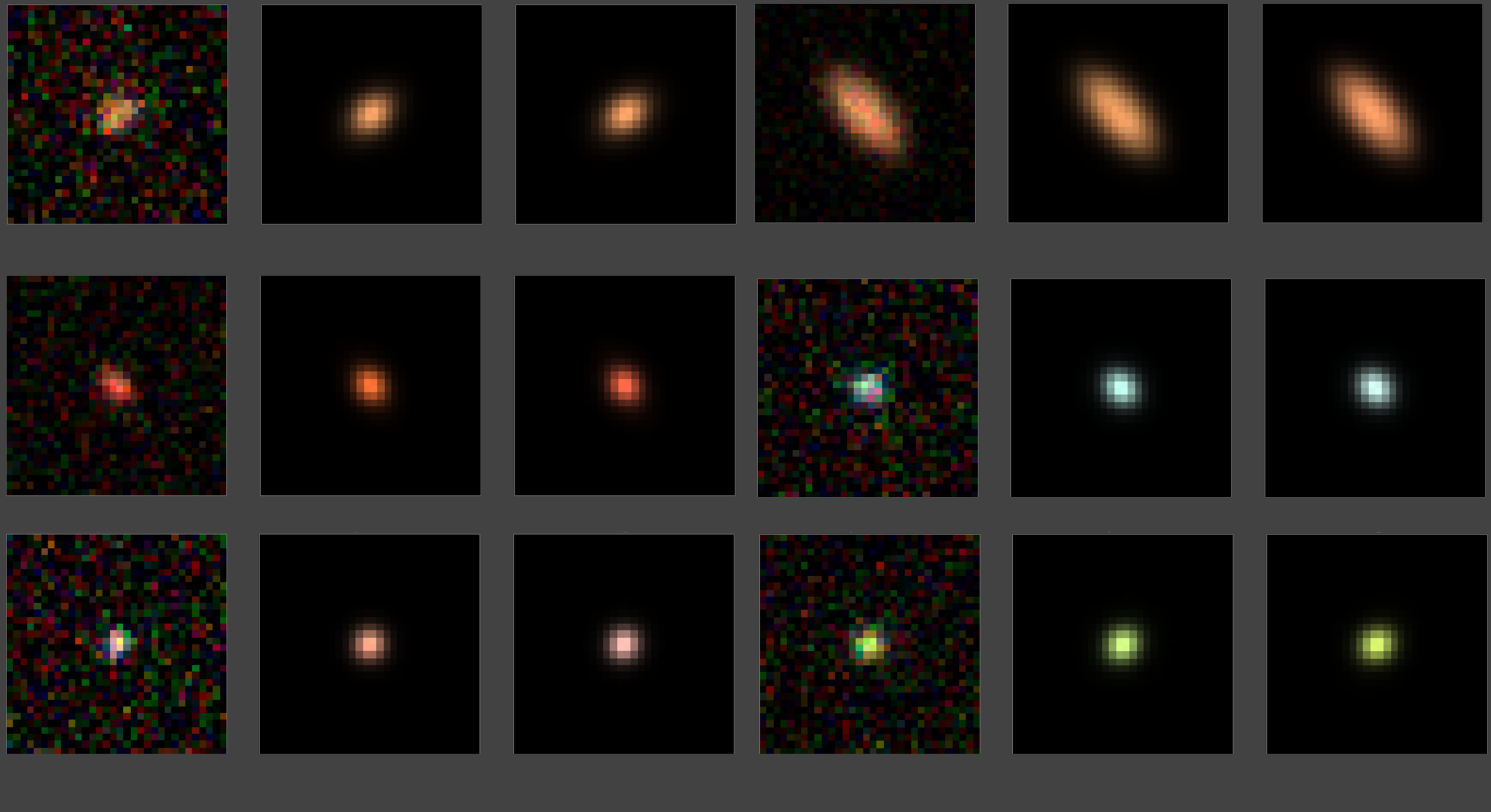
Paper under collaboration review
Arcelin et al. (2020). Deblending galaxies with variational Autoencoder: a joint multi-bands, multi-instruments Bayesian approach.

- Next step:
  Using Bayesian neural networks (TensorFlow Probability) to output ellipticities and redshift distribution from images.

Additional slides

VAE results

Few examples



Deblender results

Few examples